



Institution: Scripps Library, Miller Center at the University of Virginia

Project Title: Connecting Presidential Collections (IMLS Grant # LG-61-11-0157-11)

IMLS Award Amount=\$66,170.60, Total Project Cost=\$98,647.21

Grant Date: December 1, 2011, to November 30, 2012

Project Director: Michael Greco

Project Summary

Through its IMLS Planning Grant, the Miller Center at the University of Virginia worked to discover ways to make presidential collections more accessible. Currently there are vast amounts of presidential materials located in various presidential sites and libraries across that country. It makes it very hard for scholars, students, and presidential history buffs to know what materials are available and where to find them.

In 2010, the Miller Center co-hosted the quadrennial Presidential Sites and Libraries Conference and conducted a survey of more than 50 archivists and librarians attending the conference. The survey brought to light a central problem within the community: although there are large amounts of presidential materials available, people are simply not finding them. The reasons are complex: there are often multiple sites for a single president, it is hard for scholars and students to know what any one site has available (either physically or digitally), and there is no central resource to guide those interested on where to look or who to contact to acquire guidance. In addition, the community of presidential sites and libraries tends to be an isolated one. At the time of the survey, none of the respondents had an active partnership with another presidential site, and none had submitted data about their collections to a common descriptive catalogue.

To overcome these challenges, the Miller Center applied for an IMLS National Leadership Planning Grant with three main goals: 1. hold a Summit Meeting with experts in the field to consider the issues and possible solutions, 2. disseminate information about the project, and 3.

develop a beta website. In our grant application, we stressed collaboration and partnerships to build a resource that could benefit not only users but also organizations whose important historical resources could be discovered. The Miller Center imagined creating a centralized resource by building a website that aggregated metadata from partner collections to allow users to search across presidential materials from different institutions.

Project Process and Results

The Miller Center began its project by planning and holding a Summit Meeting that brought together a committee of experts and professionals in digital scholarship to consider the problems, examine the technological solutions available, and make recommendations on how to proceed. The Summit Meeting was held on May 17 and 18, 2012, and included sessions on short- and long-term goals, building partnerships, metadata, policies and procedures, and the front-end and back-end technologies for the beta product. The Advising Committee included archivists, librarians, documentary editors, historians, and professors in the digital humanities. (Please see attachment A, Summit Meeting Agenda, for more details.) The results of the Summit Meeting were pulled together in a summary that was shared with participants in June. (Please see attachment B, Summary of Summit Meeting.) The Miller Center also drafted Partner Policies based on the recommendations of the Summit Meeting participants. Those policies were shared with both the Summit Meeting participants and the project partners. (Please see attachment C, Policies for Partners.)

Many of the Summit Meeting participants remarked on what meaningful topics we had covered during the two days and how excited they were about our project. We were grateful for their interest and enthusiasm, and we learned an incredible amount from their varied experiences. The tone of the meeting was collegial and serious, and we appreciated the professionalism and intensity they brought to the sessions.

The second part of the planning grant was to disseminate information about the project. Although the dissemination was originally conceived as a white paper, we adapted it to conference presentations on the advice of our Advising Committee. The committee members suggested that conference or poster presentations provided the opportunity to reach more people and to have meaningful one-on-one conversations about the project. We presented the project at the Society of American Archivists Research Forum in August 2012 and at the Community Showcase at the Digital Library Federation Forum in November 2012. The [poster](#) from the DLF presentation can be found at the beta website (presidentialcollections.org). The Miller Center also maintained a project blog (presidentialconnections.wordpress.com) that followed the progress of the project and examined the issues that we encountered during the grant cycle.

During the poster and conference presentations, some of the themes that we stressed included access, collaboration, and partnership. The mission of the project is to make presidential materials more accessible. That goal includes not only helping users find materials but giving them a centralized resource so that they do not have to search multiple sites and databases to find out where their item is located. We also talked about the importance of building partnerships with presidential sites and libraries because the project was dependent on willing partners that wanted to take part and understood the benefits to their organizations and the field of presidential studies.

Beta Website

The final goal of the Planning Grant was to develop a beta website that aggregated metadata from the partners and allowed users to search across all the collections. The beta website evolved into [Connecting Presidential Collections](#) (CPC). The partner organizations whose collections made up the beta website were: the Massachusetts Historical Society, the Presidential Oral History Program at the Miller Center, the Rutherford B. Hayes Presidential Center, The Sixth Floor Museum at Dealey Plaza, the Theodore Roosevelt Center, and the Woodrow Wilson Presidential Library and Museum. The CPC site launched with the metadata for 11,349 items including photographs, letters, speeches, video clips, physical objects, pamphlets, newspaper articles, political cartoons. The beta website used Solr as the search index and Blacklight as the user interface.

The beta site included metadata from six partners. Originally we envisioned nine partners but along the way we found that some partners' collections were not ready to be included. In order to be part of CPC, the partners had to be able to pull metadata (title, description, date, author, etc.) about objects in their collection and then send that information to us in some format. We were very flexible about how they sent us the information. The ideal format was an XML file but we were also able to convert, for example, CSV documents and Excel spreadsheets. Some of the proposed partners did not have digital collections, did not have a system that allowed them to pull metadata, or did not have staff on-hand with the know-how to extract the data.

During the Summit Meeting, based on the recommendations of the Advising Committee, we made the decision to use Dublin Core as the metadata schema for the project. We chose Dublin Core because it is flexible and adaptable, and in many ways it was very appropriate for the beta site. However, Dublin Core has some limitations that we will need to consider as we move forward with a final version of the site. For example, it is not suitable for letters—there are no fields for To, From, or transcriptions. There are ways in which we can adapt Dublin Core for this purpose but it is far from perfect. One major challenge is the variety of items held in the partner collections, and no one metadata schema will be perfectly suited for such a wide variety of content types: letters, video, audio, photographs, physical objects, speeches, political

cartoons, etc. We predict, no matter what metadata schema we ultimately use, we will encounter challenges of this sort.

On the technical side, we encountered some challenges with the CPC website. First, we had some difficulties with the vendor we used to do the development work. The vendor was unable to meet our schedule and did not deliver a working configuration of Solr and Blacklight until much later than expected, which reduced the amount of time we had to make changes to the beta site. We learned that we should have had a firmer timetable for work completion—we were too flexible with the vendor and did not have firm deliverables against which we could check progress. We should have more clearly articulated what deliverables we wanted to see and when we wanted to see them. We should have been more concerned when our timetable began to slip, and we should have brought our concerns to the head of the firm more quickly than we did.

More generally, we learned a number of important lessons from the Planning Grant process. Two such lessons are that in-house knowledge and deep expertise cannot be replaced. We did not have in-house knowledge of the tools we used for the beta website and, as it turns out, our vendor did not have deep expertise. If we had known the tools, we might have recognized sooner the problems with the vendor, and if the vendor had known the tools intimately, he might not have encountered so many problems. We subsequently hired a vendor with the deep knowledge necessary to finish the work, which made a significant difference in our experience and in the quality of the final product.

We also discovered that approaching our Advisory Committee and partners with an attitude of openness created a sense of camaraderie and investment. Our relationships with both the members of the Advisory Committee and the partner organizations were positive, productive, and resulted in a better beta website. Through this project, the Miller Center is working to create meaningful relationships throughout the world of presidential sites and libraries, and we built a solid foundation for the next phases.

Evaluation Outcomes

In the Narrative of the grant application, we outlined three evaluation outcomes beyond the basic deliverables of the Summit Meeting, dissemination, and the beta product. These evaluation outcomes had mixed success. One was to establish required minimum standards for archival preservation. After discussing this item with the Advisory Committee, it was decided that this item overstepped our boundaries as a metadata aggregator. We had no interaction with the physical collections of partners, and therefore no authority to impose these archival preservation standards. We were also concerned that we would damage relationships with partners if we attempted to dictate to them how they handled their physical collections, an item clearly outside the boundaries of this project.

The second evaluation outcome was to establish basic digital standards including digitization standards, metadata standards, and best practices for creating and maintaining digital collections. This outcome proved to be more complicated than we anticipated. Although we could easily have chosen any number of digitization standards or guidelines for creating digital collections, the partners that we worked with in the planning stage had already made those decisions. They already had established digital collections and had already created their own standards. We plan to work on this outcome more fully in subsequent phases of the project because we want to have the ability to help organizations set up their digital collections and we will then need to provide them with standards and best-practice guidelines to create their digital collections.

The third evaluation outcome has two parts: a final report on the project and a detailed catalog of presidential materials. The first part, the final report, is this IMLS white paper. We would like to use this final report as a starting point for a journal article on this project and on aggregating metadata more generally, and try to have it published. As we worked on the planning grant, it became clear to us just how convoluted the world of presidential collections is, making us realize that it was all but impossible to complete the catalog with the resources allotted. In light of the importance of this catalog, we have decided to roll it into a subsequent grant application, which will afford us the resources to properly complete the task.

Future Plans

In subsequent development, we have a number of improvements we would like to make to the CPC website. Some specific examples include adding images—especially thumbnails so that users can see the object before they click through to the partner’s website—and improving the facets to narrow search results. Right now the facets are not very useful because there is not enough standardization in the partner metadata to use the facets meaningfully. For example, in the beta, the Date facet is not very useful because the metadata does not have a standard format for representing dates. In addition, for reasons we do not know at this point, some of the data for the facets is not being either indexed by Solr or displayed by Blacklight. We will need to investigate further to improve this functionality.

Two of the most important challenges we faced with this project and ones that will be encountered by other organizations that aggregate metadata are the quality of the partner metadata and the lack of standardization. The amount of metadata gathered by the partner organizations and the fields they use vary widely between organizations. We hope that as this project continues we will be able to help partner organizations improve their metadata and standardize it so that it is more uniform across organizations. We believe that, in many cases, partners will want to improve their metadata so that their items in the beta website are as useful and detailed as others. Still, many questions remain about how many resources partner

organizations will be able or willing to expend to improve their metadata. There might also be issues with the content management systems that they use and whether the systems will allow the partners to modify their metadata fields.

We would also like to *add* metadata fields across the partner organizations. Without changing the metadata that we have aggregated, we believe that we can supplement it with some fields across collections that might be helpful to users. For example, we could add a location field for each partner organization that would allow users to know where the physical object lives. Another example might be standardizing the Type field using a controlled vocabulary (such as text, image, video, audio, and physical object). This would be created as a supplemental field to improve user experience and would not overwrite any existing Type data. This kind of metadata enhancement would add meaningful facets for users, and we would happily provide this data back to the partner organizations if they were interested.

We would also be interested in running scripts across the partner metadata so that we can make some minor improvements to result in a better experience for users. To indicate inches, for example, some partners use in. and some use “. We might be able to run a script to standardize both terms to become inches, which will be more specific and meaningful. Obviously any tweaks we make to the metadata we are aggregating will have to be able to be done systematically. The amount of data we hope to eventually include into CPC means that we cannot deal with specific fields—we have to be able to apply tweaks across the whole corpus of data.

Metadata aggregation will continue to be a focus for the library community, and projects such as the Digital Library of America (DPLA) will grapple with many of the challenges we encountered during this planning grant. After talking with Emily Gore, Director of Content for the DPLA, we are excited to continue conversations with her about the challenges of metadata aggregation. We would like to learn the DPLA’s recommendations on how to structure metadata to make it more broadly available.

To continue to make presidential materials more accessible, it is necessary to make the CPC a robust and valuable resource for scholars, librarians, students, and presidential history buffs. The more collections that are added to CPC, the more people will turn to it as the first place to go when looking for presidential materials. To accomplish that goal, the Miller Center needs to add more metadata from more digital collections to CPC. There are presidential sites and libraries that have digital collections ready to be incorporated into CPC. We simply need to reach out to them and bring them on as partners on the project.

There are a number of presidential sites and libraries, however, whose collections are not yet ready for inclusion in CPC. In order to make their collections more discoverable, we need to consider ways to support them. These organizations may need tools to build digital collections

or they may need guidance in crafting metadata or they may need to be partnered with organizations that have experience in scanning fragile historical objects. We need reach out to those organizations, talk with them about their challenges, and determine the best ways to help them.

Another issue with incorporating presidential materials into CPC is making sure that we know where they are located. The world of presidential materials is chaotic, and it would be easy for us to overlook a collection whose inclusion would be valuable to the CPC. A catalog of presidential materials that identified the various collections would help overcome this issue. Although we had proposed creating a catalog for this planning grant, we quickly realized that it was an ambitious undertaking that did not have sufficient resources. The catalog will also be a valuable tool on its own as it will help scholars and librarians quickly and easily identify where to find presidential resources.

Finally, a large number of presidential materials for 18th- and 19th-century presidents exist only on microfilm, a format that is largely inaccessible. We recommend devoting research and development time to considering the most efficient ways to digitize those materials and then make them discoverable. Currently it is not easy to make digitized microfilm more than a pile of images with limited metadata. It would be valuable to consider ways to create document-level metadata for presidential papers on microfilm. If successful, that metadata could be incorporated into CPC. The Miller Center would like to continue to grow the Connecting Presidential Collections website so that it can become the first place that scholars, students, and researchers turn to when looking for presidential resources.

Web Resources

Connecting Presidential Collections beta website: <http://presidentialcollections.org/>

Connecting Presidential Collections project blog:

<http://presidentialconnections.wordpress.com/>

Connecting Presidential Collections conference poster:

<http://presidentialcollections.org/assets/imls-poster-32765911c8b59f14a1f47f7a8fe99731.png>

Attachments

Attachment A: Summit Meeting Agenda

Attachment B: Summit Meeting Summary

Attachment C: Policies for Partners